B. S.

USING SIMULATION TO DEVELOP AND VALIDATE ANALYTICAL
EMERGENCY SERVICE DEPLOYMENT MODELS

Edward J. Ignall
Peter Kolesar
Warren E. Walker

August 1975

D D C

NOV 17 1976

B

P-5463

USING SIMULATION TO DEVELOP AND VALIDATE ANALYTICAL EMERGENCY SERVICE DEPLOYMENT MODELS*
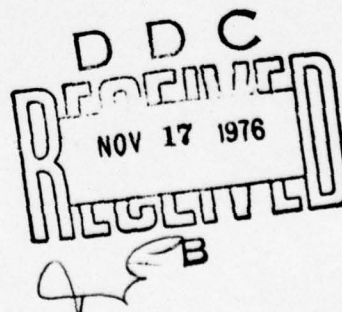
Edward J. Ignall,
Peter Kolesar
Warren E. Walker

Aug 75

44p.

P-5463

ABSTRACT

Simulation models are generally costly tools to use in systems analyses. Whenever applicable, a simple analytic model is preferable. However, in many cases, the conditions assumed by solvable analytic models do not hold in the real world. But a simulation can be used to suggest an approximate model and to determine how good an approximation an analytic model is. ~~We show how~~ simulations of New York City's fire and police operations were ~~have been~~ used to develop and validate simple analytic models that are now being used to analyze the deployment of resources in these two services.

296600

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

I. INTRODUCTION

A simulation model of a large, complex system can be very useful, but time-consuming and costly. Whenever applicable, one prefers to use a simple analytic model yielding closed-form algebraic expressions relating system inputs and outputs. However, in many cases the conditions assumed by solvable analytic models do not hold in the real world and more realistic models are too complex to solve -- hence simulation. The standard use of simulation is to answer a specific question or to describe the behavior of a system as some of its parameters are changed. In contrast, we discuss here the use of simulation to confirm that a simpler model may safely be used to describe system behavior or even to suggest the form of such a simpler model. If the analytic model provides an adequate approximation, it can be used more economically than the simulation for future analyses.[*]

Use of simulations for the development and testing of other mathematical models is analogous to experimentation by physical scientists in the development of new theory. It is this sometimes overlooked use of simulation models on which we focus in this paper.

Large-scale simulation models of fire department operations and police patrol activities have been developed at The New York City-Rand Institute. Although, at present, the fire simulation models one borough of New York City and the police simulation models several New York City police precincts, insights obtained from each of the simulations have been used to produce and verify several analytic models having general applicability throughout New York City and elsewhere.

---

[*]We assume in this discussion that the simulation in question is an adequate (valid) representation of reality. The particular simulation models discussed below have already been tested for "face" validity [14].

We discuss the validation of four such models in this paper:

(1) A model for analyzing police patrol car allocation problems, where the simulation and analytic models were constructed in parallel. One of the chief reasons for building the simulation was to determine how well and under what conditions the analytic queueing model agreed with it (and thus with the real world).

(2) A model for estimating fire company response distances, where the analytic model was developed well after the simulation was written and special simulation runs were made to confirm its validity.

(3) A model for predicting the number of fire companies dispatched to an alarm, where the analytic model was suggested (and verified) by an analysis of simulation runs that had been made years earlier for other purposes.

(4) A model for estimating the number of fire companies that will be busy at alarms, where the model was developed before the simulation was written but was verified several years later using the results of simulations that were run before the validation effort began.

II. A MODEL FOR ALLOCATING POLICE PATROL RESOURCES

In [13]* a queueing model is proposed to represent the patrol activities
of a police command. A patrol car is dispatched immediately to answer a
call for service if one is available; otherwise, the call is queued at the
dispatching center. Queued calls are assigned a priority, and high priority
calls are served before those of lower priority. What is desired is a way
of relating queueing delays and car availability to N, the number of cars
assigned to the command, so that N can be chosen rationally.

The queueing model is the simple M/M/N priority model of Cobham [6],
which assumes that calls arrive according to a stationary Poisson process,
that service times are independent and exponentially distributed, and that
each call is served by a single patrol car. These conditions are not all
satisfied in the operating environment of the New York City Police Depart-
ment (NYPD). While call arrivals for any short interval are approximately
Poisson distributed, call rates, even during a single 8-hour tour of duty,
are not constant. Service times are not exponentially distributed, and in-
clude the time required for a car to travel to an incident, which depends
on the number of cars available to dispatch. Moreover, a call may be
served by more than one patrol car.

As a result, although we wanted to use the stationary queueing predic-
tions of the simple M/M/N model to analyze deployment options for the NYPD,
we first had to verify that, despite the above-mentioned variations from the
model, it still produced predictions of sufficient accuracy. To make ap-
propriate tests, we wrote a detailed police patrol simulation of a single
police precinct [11]. The simulation included the complexities mentioned
above, as well as others, and used actual call histories in the precinct for
arrivals and service times. We compared simulation results to those obtained
from the queueing model with the same average call rate and the same average

---

*Numbers in square brackets indicate references listed at the end of
this paper.

service time. The results, described below, were close enough to give us
and the Police Department confidence that the queueing model could be used
instead of the simulation model to analyze some important deployment
problems.

Based on the call rate, average service time, and number of servers,
the queueing model gives the probability distribution of the number of calls
being serviced and the number waiting to be dispatched. From these probabil-
ities a great deal of information about the performance of the system can
be obtained. For example, suppose N patrol cars are on duty and let P_j
be the probability that there are j calls in the system (being served and
waiting to be served). Then, two of the quantities that can be calculated
are:

The probability that all N patrol cars are busy (q_N):

$$q_N = \sum_{j=N}^{\infty} P_j. \tag{1}$$

The average time a call will spend in queue before being
dispatched ($\overline{D}$):

$$\overline{D} = 1/\mu \sum_{j=N}^{\infty} (j + 1 - N) P_j, \tag{2}$$

where $1/\mu$ is the average service time.

To test the usefulness of this model we calculated these quantities
as functions of N and compared them to results obtained from the simulation
model. The 71st Precinct in Brooklyn was chosen for study because a par-
ticularly rich set of data on its operations was available. We analyzed re-
cords of all calls for service received within the precinct during the
months of August and September 1972. The calls were aggregated by their
time of occurrence into the three shifts or "tours" worked by the policemen:

Tour 1, midnight to 8 a.m.; Tour 2, 8 a.m. to 4 p.m.; Tour 3, 4 p.m. to midnight. The average service time was approximately the same for all tours. The queueing model was used to analyze conditions for Tours 1 and 3 with different numbers of cars on duty. The simulation was run for the same values of N, using as input a historical stream of calls for service for a given tour. We used the actual time each call was received and its actual service time, location, and priority. (The input stream for the simulation of a given tour was prepared from computerized records maintained by the NYPD by concatenating all of the calls received during that tour for the months of July and August 1972. For example, when simulating tour 3, the last call before midnight on one day would be followed by the first call after 4 p.m. on the following day.)

A comparison of the results from the simulation and queueing models is given in Figs. 1 and 2. Figure 1 shows the percent of time that all patrol cars are busy. The results are remarkably similar, with the queueing predictions being consistently slightly lower than the simulation results. (We predicted this difference because the simulation makes multiple car dispatches while the queueing model assumes that one car is sent to each call.) Figure 2 plots the average queueing delay as a function of the number of patrol cars on duty. The results, again, are quite close. Because of the type of applications we had in mind, we were particularly interested in finding out whether the queueing model would predict the "elbow" in these functions -- the point at which the curves begin to rise steeply and performance begins to degrade badly. It appears to do so.

The Police Department accepted the fact that the queueing model does represent a reasonable approximation to the dispatching and service activities

-6-

RESULTS OF POLICE PATROL SIMULATIONS TO VERIFY QUEUEING MODEL

● Simulation Results
○ Predictions of Queueing Model

DATA: 71st. Precinct
July - August 1972

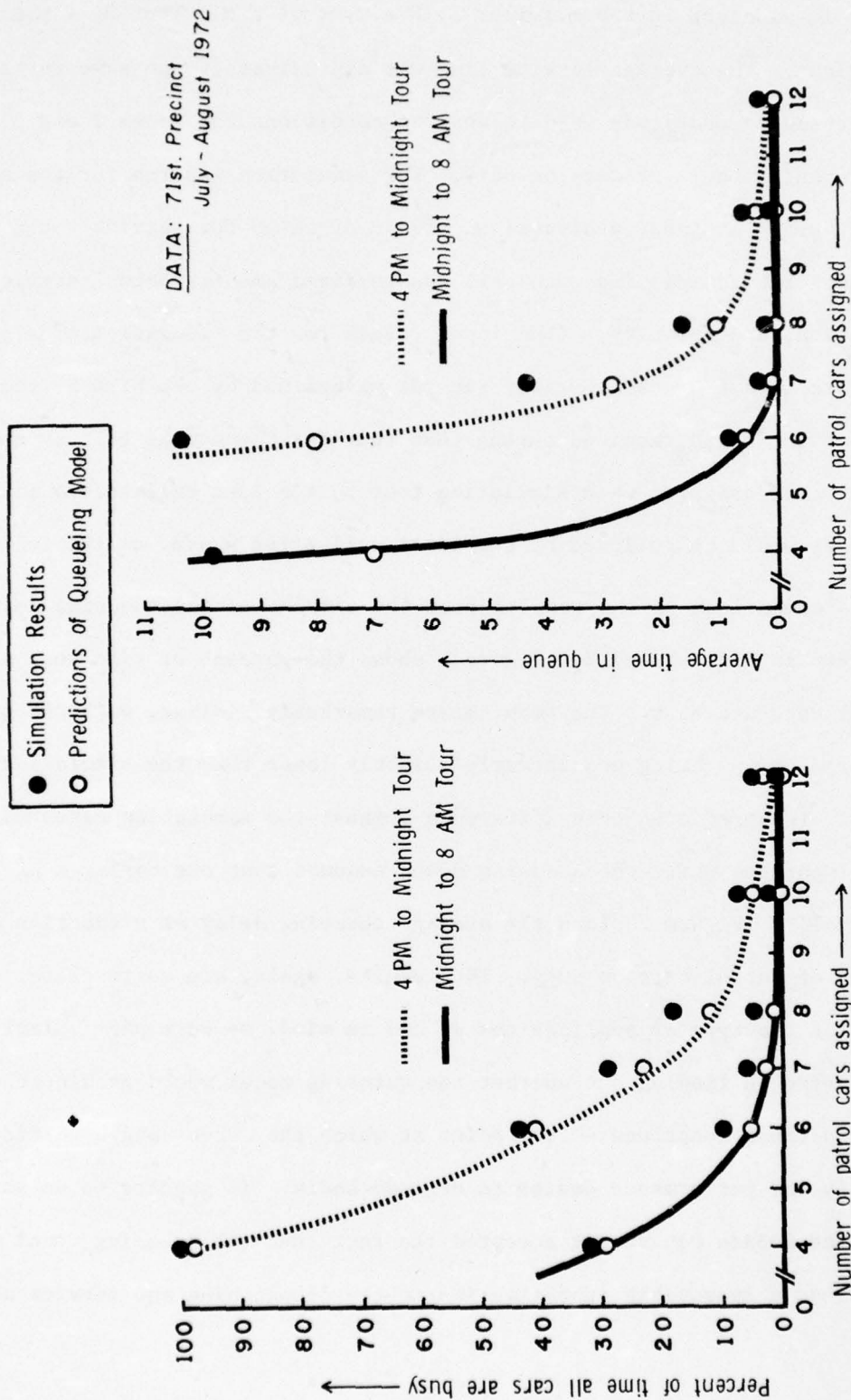FIGURE 1: Unavailability of patrol cars vs. number of cars assigned

FIGURE 2: Delay in queue (in minutes) vs. number of patrol cars assigned

of the patrol force. The model has been imbedded in a computer program
called the Patrol Car Allocation Model [5], which the Police Department has
begun to use as an aid in determining the number of patrol cars to assign to
duty during each tour in each police precinct.

III. A MODEL FOR ESTIMATING FIRE COMPANY RESPONSE DISTANCES

In [10] Kolesar and Blum derive an inverse square-root relationship between the average distance traveled by fire companies responding to calls for service and the number of locations in the region from which they respond. The relationship was derived rigorously under idealized conditions: an infinitely large region in which the units are located either uniformly on a grid or purely at random, and in which calls for service are distributed homogeneously in space, while emergency vehicles travel along simple response paths. However, to have practical usefulness, it was important to show that the relationship provided a reasonable approximation to actual average response distances under more realistic conditions.

An existing simulation of New York City fire-fighting operations was used to test the validity of the model for fire company responses. The simulation program is described in [1], details of its design are given in [2], and some of its uses for policy analysis are described in [3]. Two hypothesized relationships were to be tested by simulation:

- The expected response distance, ED(N), of the closest available unit to a fire alarm occurring, when there are N _available_ companies in a region of area A, is given by

$$ED(N) = k\sqrt{A/N}, \qquad (3)$$

  where k is a constant of proportionality that depends on the street configuration and the manner in which units are distributed throughout the region.

- If there are n companies located in a region of area A and if, on the average, b are busy, then the average first company response distance to all calls for service in the region is approximately given by

$$\overline{D}(n) = k\sqrt{A/(n - b)}, \qquad (4)$$

where n-b is the long-run average of N and $\bar{D}(n)$ is the long-run average of ED(N). Since n is a major policy variable under management's control, this relationship is of more general interest than (3).

Verifying these relationships by gathering empirical data from Fire Department operations would be an extremely difficult task. In fact, verifying (4) would require the Department to vary the number of units operating in an experimental region at different times—an unthinkable procedure, especially if the changes meant using so few companies that lives and property were endangered. Instead, by means of simulation, these tests could be made safely and economically without any modifications to actual Fire Department operations.

To validate and test the model, seven simulations were run modeling the fire-fighting operations in the Bronx, one borough of New York City. The conditions simulated varied over a broad range. Alarm rates were varied from 5 to 30 alarms per hour. In each case, the alarms were distributed so that the probabilities of occurrence among 358 locations were as they were in 1968. The number of active ladder companies was varied from 12 to 31, while the number of engine companies was held constant at 37 for all runs. The engine locations were those actually being used in the Bronx in 1971. The ladder locations in the 12-, 20-, and 24-ladder cases were subsets of the 27 actual ladder locations in the Bronx in 1971. In the 31-ladder case, 4 additional ladder locations were added in the south Bronx in carefully chosen places. In each case an extended time period was simulated during which the alarm rate and number of active units was unchanged. The simulation durations were chosen so that, in each case, about 3,500 alarms were handled. This sample size was selected after statistical analysis of the random variation in simulation output statistics. The results produced should be interpreted as estimates of performance of "steady state" behavior under the conditions simulated. Table 1 gives a brief listing of the simulations carried out.

Table 1

LIST OF CONDITIONS SIMULATED

Simulation Number	Alarm Rate (Alarms/Hour)	Number of active Ladders	Number of active Engines
1	30	31	37
2	5	31	37
3	5	24	37
4	10	20	37
5	30	20	37
6	5	12	37
7	10	12	37

RELATIONSHIP (3)

First, we consider the validation of relationship (3) between ED(N) and N.
The simulation program recorded the response distance and the number of com-
panies available at the instant of dispatch for each alarm. These data
were accumulated separately for two regions of the Bronx; these were, ap-
proximately, the south Bronx (a small region with a high incidence of alarms)
and the rest of the Bronx (called "the north Bronx"). The data were collected
for the closest engines and ladders to each alarm, as well as for the second
and third closest units for those alarms to which such units were dispatched.

In order to analyze these data and determine if the square-root model
was appropriate, we graphed average response distance vs. the number of
units available for each individual set of data. (By an individual set of
data we mean, for example, data for second closest ladders in the north Bronx
from the simulation run with 12 active ladder companies and an alarm rate of
10 alarms per hour.) In addition, for each set of data we used the method
of least squares to determine the parameters (α, β, and k) of two response
distance models:

$$D = k(A/N)^{-1/2} \qquad (5)$$
$$D = \alpha(A/N)^{\beta} \qquad (6)$$

We were concerned with how well these models fit the data. If square-
root relations indeed hold, relations (5) and (6) should both fit well,
estimates of β should be "close" to $-1/2$, and k and α should be nearly the
same. Measuring how close β is to -0.5 is not straightforward, since the
simulation data do not satisfy the conditions requisite for classical
statistical analysis. For example, the observations are not independent, and
the square-root relation itself implies unequal variances. Examination of
the sum of squared errors from our regressions indicates, of course, that (6)
fits better than (5), but the difference between the models is small, and

an "eyeball" check of the graphs in each case revealed little difference
between (5) and (6).

This analysis was done for closest and second closest engines and
ladders, for the north and south Bronx, for each simulation--a total of
28 cases.   We saw a general consistency between the square-root law
parameters for engines and ladders and between results from the north and
and south Bronx, so we repeated the analysis with the data grouped from
various simulations and for engines and ladders. This grouping yields
parameter estimates in which slight differences due to the geography of
the regions and to the particular company locations are averaged out, and
for this reason should be more appropriate for use in other regions of
the City.   Figures 3 and 4 display some of these results, and Table 2
provides a summary. We note that the R^2 values for the square-root model
are quite high (see Table 2), that the graphs show a close correspondence
between the fitted functions and the data, and that the values for β are
all very close to -0.5.

Data from previous simulations, which had been made for other purposes,
were analyzed in the same way and yielded similar results.[*]   On the basis
of all these results, we concluded that the data confirmed the validity of the
square-root law between average response distance and the number of units
available at the moment of dispatch.

RELATIONSHIP (4)

We now turn attention to validation of relationship (4) between long-run
average response distance, $\bar{D}$, and n, and the number of companies assigned
to the region. The data and analysis just discussed indicate that the

_____

[*]Some of these results appear in [10].

Figure 3. Simulated average response distances vs. the number of available companies (closest engine and ladder companies in the south Bronx).

Figure 4. Simulated average response distance vs. the number of available companies (second closest engine and ladder companies in the south Bronx).

Table 2

FITS OF SQUARE-ROOT MODEL TO GROUPED DATA

Grouping	Parameters of Exponential Model		Parameters of Square-Root Model	
	α	β	k	R^2
Closest Engines and Ladders				
North Bronx	0.57	-0.47	0.60	0.93
South Bronx	0.57	-0.50	0.57	0.97
Second Closest Engines and Ladders				
North Bronx	1.23	-0.51	1.22	0.73
South Bronx	1.13	-0.48	1.16	0.88

Exponential model: $D = \alpha(A/N)^\beta$.

Square root model: $D = k\sqrt{A/N}$.

R^2 = Sample correlation coefficient for square-root model.

square-root model describes the relationship between average response distance and the number of units available when an alarm occurs. But this does not assure that a square-root law describes the relationship between long-run average response distance and the average number of companies available to respond to an alarm. On the contrary, if the square-root law holds for the former, it cannot hold exactly for the latter since the inverse square-root function is convex, and for a convex function $f(\cdot)$ of a random variable X, $Ef(X) > f(EX)$ (Jensen's inequality). The question of practical interest, however, is whether or not the square-root model provides an adequate approximation.

Figure 5 displays the simulated long-run average response distance for closest ladders versus average numbers of ladder units available for south Bronx ladders. The data were generated by the same set of seven simulations described in Table 1. In this case each of the plotted points represents the results of an entire simulation run. In addition to the simulation data, we have also plotted regression fits of the functions

$$\overline{D} = k/\sqrt{\text{Average number of available ladders}},$$
$$\overline{D} = \alpha(\text{Average number of available ladders})^{\beta}.$$

As before, we are interested in how well these functions fit the data, whether β is close to $-1/2$, and whether α is close to k. We are also interested in the consistency of the results across the regions. As the figure shows, the fits are good and the two functions nearly coincide. Table 3 summarizes the results of several regressions we did. (Other regressions—not shown here—were done of the standard deviation of response distance versus the average number of companies available. The theory indicates that this relationship should also be an inverse square-root function, and analysis of the simulation data supports this hypothesis.)
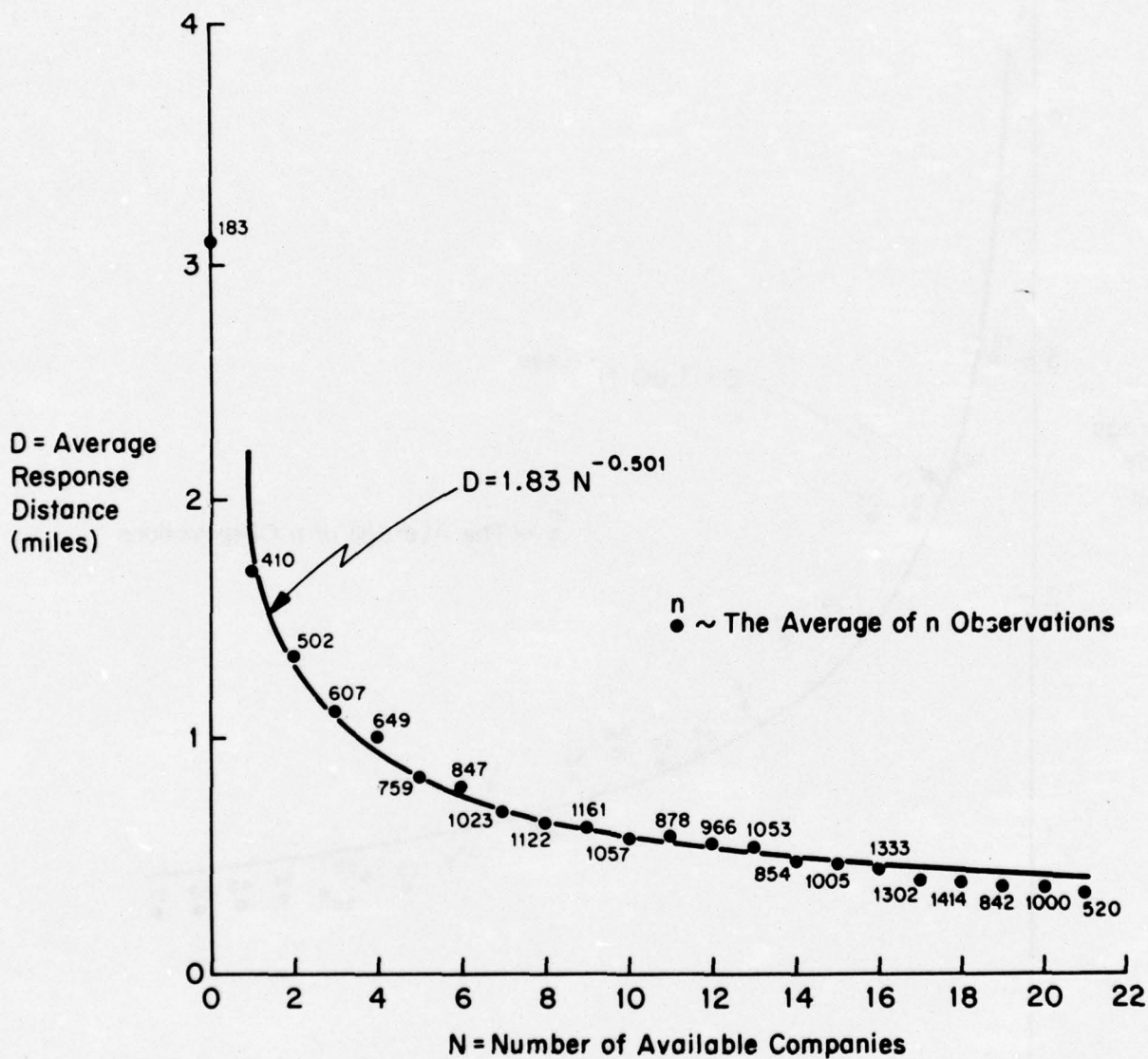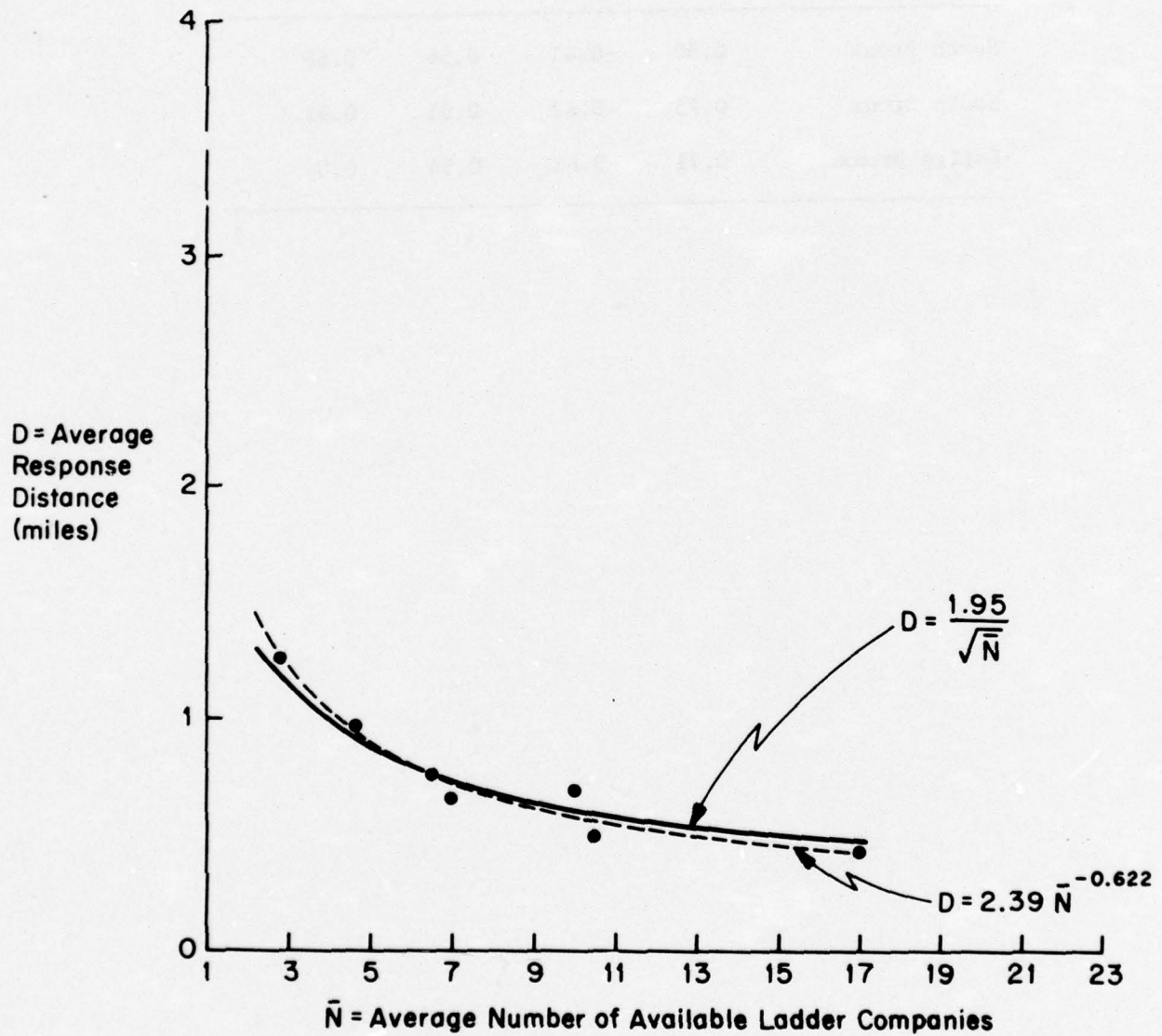
Figure 5. Simulated long-run average response distance vs. the average
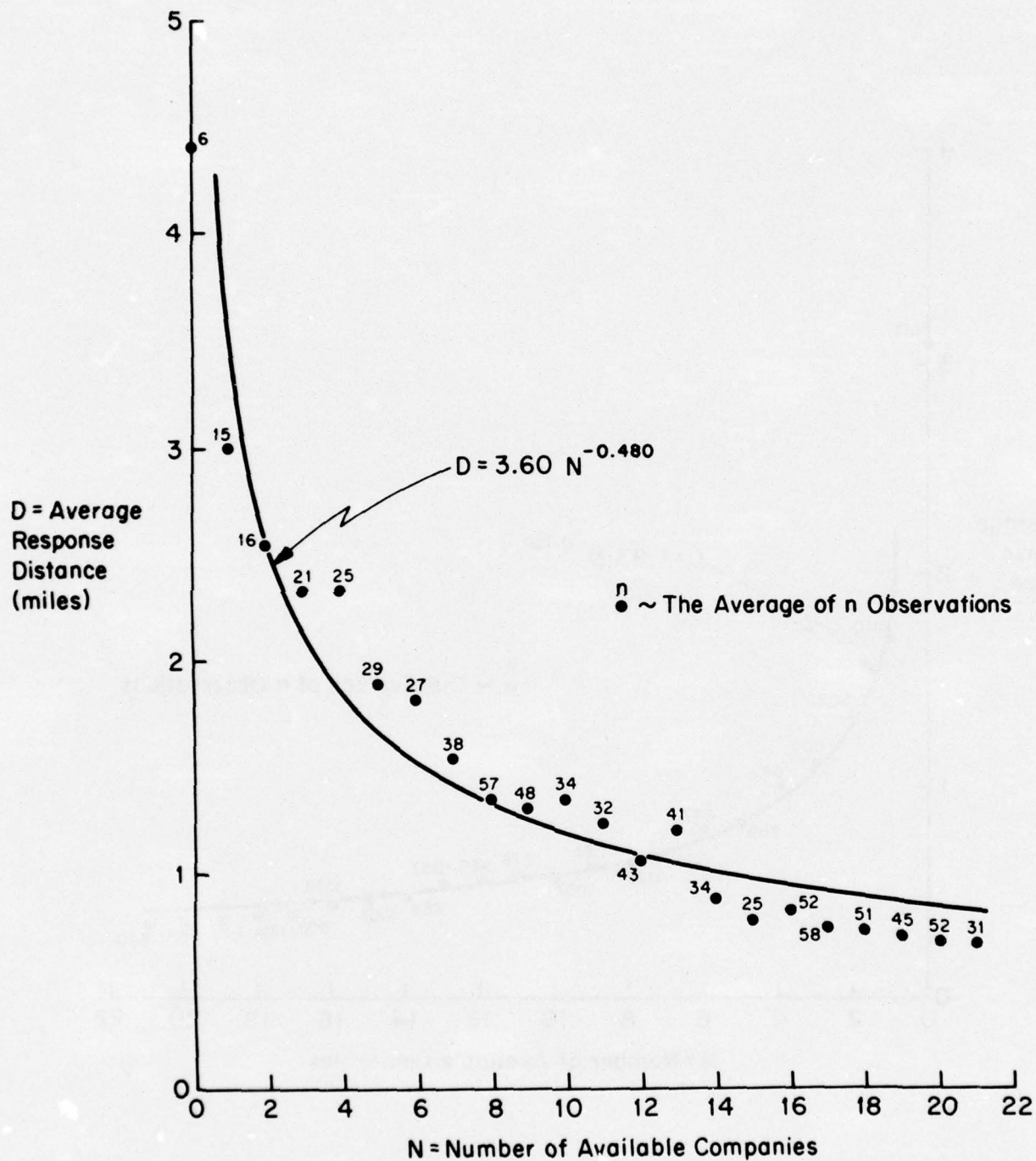number of available companies (closest ladder companies in
the south Bronx).

Table 3

RESULTS OF REGRESSIONS OF LONG-RUN AVERAGE RESPONSE
DISTANCES FROM BRONX SIMULATIONS

Region	α	β	k	R^2
North Bronx	0.50	−0.41	0.56	0.68
South Bronx	0.75	−0.62	0.61	0.92
Entire Bronx	0.71	−0.62	0.54	0.86

USES OF THE SQUARE-ROOT MODEL

The distance that responding fire units travel to reach fire alarms is an important measure of the service provided by the Fire Department. The ability to predict the average response distance in a region as a function of alarm rate, the number of units assigned to the region, and other measurable parameters of the region, permits allocation policies to be evaluated quickly without the use of simulation [10]. Among the many questions that the Fire Department of New York has already used the square-root model to answer are:

- What will be the effect on response distance of removing a company from a region?

- How should the number of units on duty be varied over the day (as the alarm rate varies) to maintain a given average response distance in a region?

- How many fire units will be required in the future under projected alarm rates to maintain desired average response distances?

A SQUARE-ROOT MODEL FOR POLICE PATROL CARS

The average distance traveled by police patrol cars responding to calls for service is also of interest. The main difference in travel characteristics between fire companies and patrol cars is that patrol cars move about when not servicing a call, while fire companies are generally stationed in fixed locations. Analytically, this means that in the case of patrol cars, the estimation of response distance involves computing the distance between two random points—the location of the car and of the incident. Larson [13] considers this problem in some detail, and shows that different versions of a square-root model for distances apply depending on the

circumstances.

As part of a study of the possible benefits to the NYPD of automatic devices for estimating the location of patrol cars, we carried out simulation experiments of patrol car movement using The New York City-Rand Institute's patrol car simulation model [11]. Figure 6 presents results from these experiments that show the validity of a square-root model for average patrol car response distances of the same form given in relation (4) for fire companies.

Results of two series of simulation experiments are presented. In one series we imitated the existing dispatching system (called SPRINT), which uses imperfect information regarding a car's location—the system makes assignments on the basis of the car's probable location. In the other series we imitated the performance of an automatic car locator system, which supplies the dispatcher with the exact position of the car. It appears from the data that a square-root model holds for each system. (The results also show that the improvement due to the car locator system is small, which is what we had expected.)
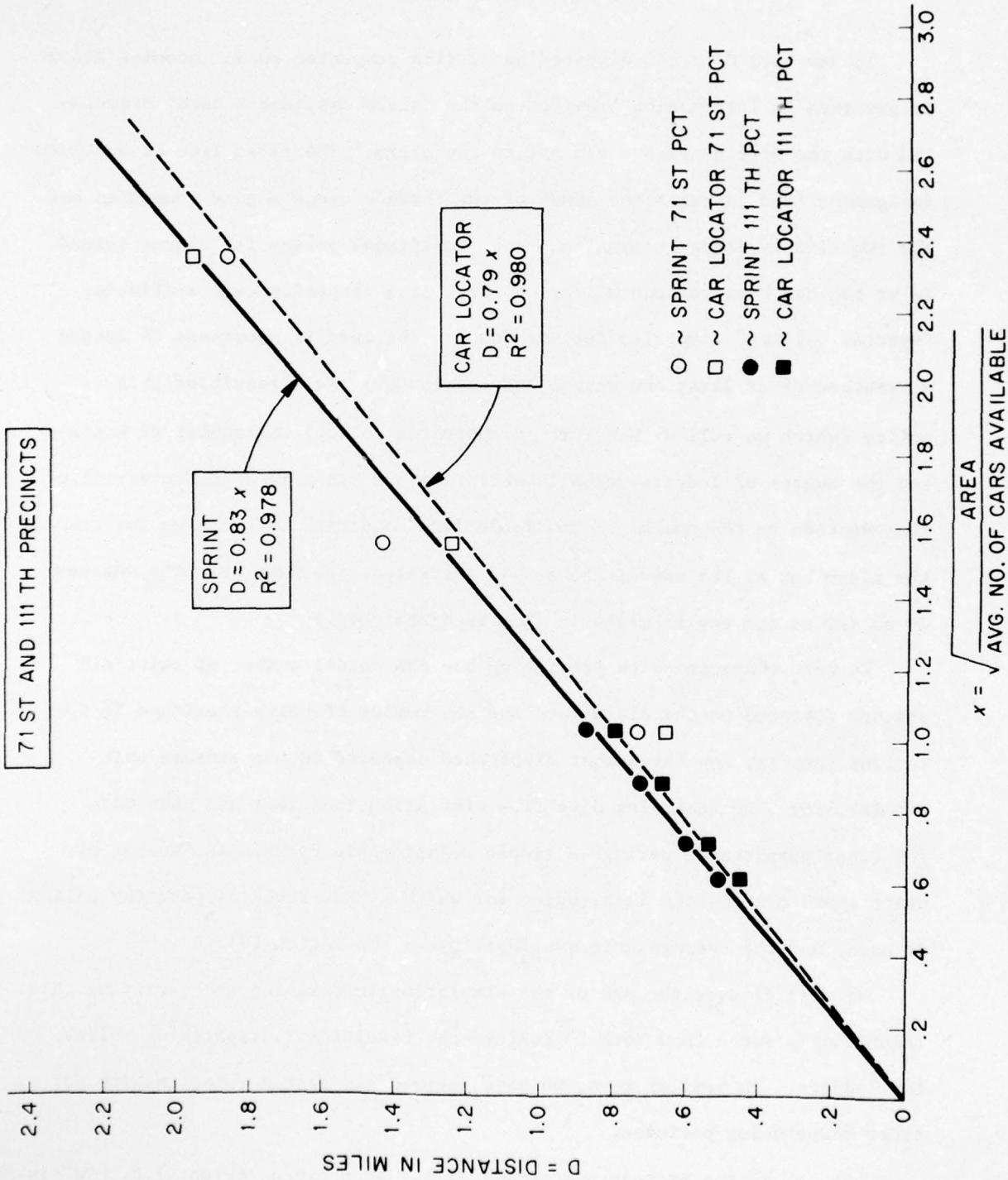
Figure 6. Simulated average response distance as a function of the square-root of the average number of patrol cars available.

IV. A MODEL FOR PREDICTING THE NUMBER OF UNITS
SENT TO A FIRE ALARM

In New York City the dispatching of fire companies to an incoming alarm is governed by information provided on the "alarm assignment card" associated with the fire alarm box closest to the alarm. The first line of an alarm assignment card contains the names of the three closest engine companies and the two closest ladder companies. The traditional policy for alarms turned in by box had been to send whichever first line companies were available, "special calling" companies further down on the card if necessary to assure a response of at least one engine and one ladder. As a result of this policy (which we call a "New York" dispatching policy) the number of engines and the number of ladders actually sent to a box alarm is a random variable that depends on the number of available fire companies in the area surrounding the alarm box at the moment the box is activated. As many as three engines or as few as one engine might in fact be dispatched.)

We were concerned with predicting how the actual number of units dispatched depended on the alarm rate and the number of units stationed in the region; that is, how the number dispatched depended on the average unit availability. By analyzing some fire simulation runs that had been made for other purposes we derived a simple relationship between the number of units sent to incidents in a region for which a "New York" dispatching policy is used, and the average unit availability in the region [8].

We will discuss the use of the simulation in deriving and verifying this relationship for a "New York 2" policy—the traditional dispatching policy for ladders. In similar ways, we have derived and tested relationships for other dispatching policies.

Let "a" be the average unit availability in a given region; i.e., "a" is the average fraction of the time a unit is available. Define $P(n, a)$ as the probability that n units are dispatched to an incident at which the New York

2 policy is used during a time period in which the average availability is a.

A good fit to the simulation data was found by using:

$$\begin{cases} P(2, a) = a^2 \\ P(1, a) = 1 - a^2 \end{cases} \tag{7}$$

This relationship would be true if

(1) The average availability were the same for every unit in ·the region;

(2) The event that any particular unit is available were independent of the status of all other units.

Neither of these is true, yet the relationship appears to be a good approximation. In fact, the validity of the relationship (7) was discovered in the course of attempting to see how poor a^2 was as an estimate of $P(2, a)$.

The relationship was developed from a set of simulation runs that had originally been made to test the effects at different alarm rates of both adding new companies to an area and modifying the dispatching policy (see [3]). As mentioned before, the simulation models the Bronx and, for data collection purposes, it was partitioned into two regions: the south Bronx, a very high activity region (which we shall call region 2), and the rest of the Bronx (region 1). Nine simulation runs were analyzed in which four different alarm rates and five different allocations of fire companies were used.

For each simulation run we calculated, for each ladder company, the fraction of time it was available. We then obtained the average of these availabilities over all the ladder companies in the region and called this average a. For all incidents in the region for which the response policy

was New York 2 we found $P(2, a)$ the proportion that received two ladders as their initial dispatch (these data had been part of the normal output from all simulation runs). The results in Table 4 show that $P(2, a)$ was very close to a^2. In particular, in the first simulation run, based on an observed availability in region 2 of .898 we would predict by relationship (7) that two ladders would be sent to 80.6 percent of all box alarms. In fact, two ladders were sent to 80.7 percent of the box alarms.

The relationship was then validated by analyzing the results on a different set of simulations that had also been run previously but had not been examined during the derivation of the relationship. The original objective of these runs had been to study the effects of matching the number of fire engines on duty more closely to the time-varying alarm rate. Results of four simulations were analyzed and compared to the results predicted by the relationship (see Table 5). On the basis of this comparison we were able to conclude that the relationship we had derived did provide a useful approximation to the actual field dispatching behavior (especially in the high alarm rate region—region 2). We have, therefore, begun to use this relationship (and others so derived for different dispatching policies) instead of the simulation to analyze the effects of various Fire Department deployment options on the number of units dispatched.

Table 4

DERIVATION OF NEW YORK 2 RELATIONSHIP

Observed Availability	Region 1 Observed and predicted (from availability) percent of NY2 alarms receiving the indicated number of ladders (observed/predicted)		Observed Availability	Region 2 Observed and predicted (from availability) percent of NY2 alarms receiving the indicated number of ladders (observed/predicted)	
	One ladder	Two ladders		One ladder	Two ladders
.952	13.2/9.3	86.8/90.7	.898	19.3/19.4	80.7/80.6
.951	21.8/9.6	78.2/90.4	.882	24.7/22.7	75.3/77.9
.877	24.3/23.1	75.7/76.9	.769	35.1/40.9	64.9/59.1
.875	25.0/23.3	75.0/76.7	.752	41.9/43.4	58.1/56.6
.871	26.5/24.1	73.5/75.9	.739	45.8/45.4	54.3/54.6
.869	27.5/24.4	72.5/75.6	.718	52.5/48.5	47.5/51.6
.776	47.3/39.7	52.7/60.3	.555	72.3/69.2	27.7/30.8
.666	58.7/55.7	41.3/44.3	.466	72.7/78.3	27.3/21.7
.623	66.2/61.2	33.8/38.8	.379	85.1/85.6	14.9/14.4

Table 5

VERIFICATION OF NEW YORK 2 RELATIONSHIP

	Region 1			Region 2	
Observed Availability	Observed and predicted (from availability) percent of NY2 alarms receiving the indicated number of ladders (observed/predicted)		Observed Availability	Observed and predicted (from availability) percent of NY2 alarms receiving the indicated number of ladders (observed/predicted)	
	One ladder	Two ladders		One ladder	Two ladders
.945	16.2/10.8	83.8/89.2	.849	27.8/27.9	72.2/72.1
.882	29.8/22.2	70.2/77.8	.705	49.8/50.3	50.2/49.7
.948	16.2/10.2	83.8/89.8	.847	28.1/28.3	71.9/71.7
.888	30.7/21.1	69.3/78.9	.699	50.7/51.1	49.3/48.9
.738	57.5/45.6	42.5/54.4	.494	75.5/75.6	24.5/24.4
.745	55.0/44.5	45.0/55.5	.485	75.8/76.4	24.2/23.6
.689	61.9/52.6	38.1/47.4	.469	86.1/78.1	13.9/21.9

V. A MODEL FOR ESTIMATING THE NUMBER OF FIRE COMPANIES
WORKING AT ALARMS

In [4] Chaiken proposed a queueing model of the number of fire companies busy working at alarms. The model gives the long-run probability distribution of the number of busy companies. Its assumptions are much closer to reality than those of the models in Sections II and III—nonexponential service times are allowed and there are no assumptions about the geographical distribution of alarms or companies. Therefore, the model had been used for planning purposes in New York City before we compared its predictions with simulation results. We discuss the model and its assumptions and how we validated it using results from simulation experiments carried out several years earlier for other purposes.

The model is as follows: Consider a region of a city in which alarms of different types are generated according to independent Poisson processes. In response to each alarm, a number of units are dispatched at once, and then additional units may be required later as the service progresses. Later, the units are released from service one at a time or in groups until finally the service is complete. Assuming that the alarm and service rates are not varying with time, the model gives the distribution of the number of busy units in the steady-state. The model assumes that infinitely many units are available, since fire departments will generally send as many units as are required, even if it is necessary to enlist the assistance of nearby cities to obtain a sufficient number. The model assumes that the service times for each alarm are independent of the number of units already assigned to previous alarms. The alarms are separated into *types* according to the number of units required to serve the alarm and the distributions of the service times for each stage of service. A *stage* of service consists of a period of time when a fixed number of units are committed to serving the

alarm. Thus, a new stage begins whenever an additional unit is dispatched or a unit completes service. Permitting some stages to have zero duration allows for the possibility that units may be dispatched or released in groups of two or more.

The stages of service are statistically independent, and the number of units engaged increases monotonically to the maximum and then decreases monotonically to zero. Thus, if a maximum of n units is to be assigned to the alarm, there are 2n-1 independent stages of service, which are numbered 1,2,..., 2n-1. If j<n, the number of units busy in the jth stage is j; if j>n, the number of units busy in the jth stage is 2n-j.

To illustrate the stages, consider one type of fire alarm, which is small enough to be extinguished by the men on two fire engines. Suppose that initially three fire engines are dispatched to the scene, so that the first two stages (with one unit busy and two units busy) have zero duration, and stage 3 begins when the alarm arrives. After a period of time, it is determined that two engines are adequate to fight the fire, and the third is released. At this time service enters stage 4, with two units busy. When the fire is extinguished, one of the two remaining engines is released, while the other remains for some follow-up activities, called "overhaul," and this constitutes stage 5.

Since the units perform certain activities together as a group, it is not appropriate to think of the service times of the various units as being independent of each other. However, since the stages constitute distinguishable activities (stage 3 = dispatch, stage 4 = extinguishment, stage 5 = overhaul), it is possible to select the alarm types in such a way that independence of the stages becomes a satisfactory approximation.

This model is a generalization of Erlang's formulas [7], which were proved valid for the case of an infinite-server queue and general service times by Khintchine [9].

Chaiken's main result can be summarized as follows:

Suppose the input to an infinite server queue consists of ℓ independent Poisson processes having rates $\lambda_1, \ldots, \lambda_\ell$. Suppose further that the process i, with rate λ_i, requires $2n_i-1$ independent stages of service, such that the jth stage has an arbitrarily distributed holding time with finite, non-zero mean $T_i(j)$ and requires $n_i - |n_i - j|$ servers, $j=1, \ldots, 2n_i-1$. Then the steady-state probability that m servers are busy is

$$P(m) = e^{-r}, \text{ for } m=0$$

and

$$P(m) = e^{-r} \left[\rho(m) + \frac{1}{2!} \rho^{*2}(m) + \ldots + \frac{1}{m!} \rho^{*m}(m) \right], \text{ for } m > 0.$$

Here $\rho(j) = \sum_{i=1}^{\ell} \rho_i(j)$, with

$$\rho_i(j) = \begin{cases} \lambda_i \left(T_i(j) + T_i(2n_i-j) \right) & \text{if } j=1, \ldots, n_i-1 \\ \lambda_i T_i(n_i) & \text{if } j=n_i \\ 0 & \text{otherwise} \end{cases},$$

$r = \sum_j \rho(j)$, and $\rho^{*k}(m)$ denotes the k-fold convolution of $\rho(m)$:

$$\rho^{*k}(m) = \sum_{j=0}^{m} \rho(m-j) \, \rho^{*(k-1)}(j).$$

These equations can be used to decide how many emergency units to locate in any specified geographical region. One calculates the probability that more than n units will be busy at one time. Then one assigns enough units to the region so that this probability does not exceed a certain threshold. This method has been used by the New York City Fire Department.

VALIDATING THE MODEL

To test the model with empirical data would be very difficult, if not impossible. An important problem in this regard is that the model gives steady-state results for a system with a stationary alarm rate, yet the actual alarm rate varies throughout the day. We decided to test the queueing model by comparing its results to results derived from the simulation model of fire-fighting operations in the Bronx. As we have pointed out, the simulation already mentioned is a more detailed and accurate model of fire-fighting operations. It incorporates several characteristics of actual operations not treated in the queueing model. For example, in the simulation, service times for fires are to some extent state dependent because travel times and the number of units actually dispatched depend on the number available when the alarm occurs. In addition, there are a finite number of units in the simulation while the queueing model assumes an infinite number. Units can be dispatched across region boundaries, and there is a "relocation" procedure that will temporarily reassign units to firehouses other than their home houses if protection in a region is too low [12].

The simulation has a set of incident types based on the number of units required by the incident and their service times. We were therefore able to designate the stages and stage durations of the incidents according to the structure of the queueing model. The simulation had been constructed years before we carried out this validation exercise. It was not designed with this test in mind and its printed output reports did not include the data we needed for this test. But, we were able to use the results of simulations run several years before this validation test because an output tape had been created from each simulation run that recorded the status of every fire-fighting unit at 15 minute intervals of simulated time. From these tapes we were able to create histograms of the number of busy units.

The data were aggregated in several ways. There are two distinct types of fire-fighting units, engine companies and ladder companies. They perform different but complementary functions at fires and so we wanted to analyze their availability separately. There are also two regions of the Bronx that were of interest for planning purposes--the south Bronx, a region of high fire incidence, and the north Bronx, a lower fire incidence area. We used data from simulations with the overall alarm rate in the Bronx varying from 5 alarms per hour (a low-rate, typical of late nights) to 30 alarms per hour (a very high rate, which was an extrapolation to future summer evenings). So we wound up with 24 sets of data (2 types of units, engines and ladders; 3 regions, north, south, and the entire Bronx; and 4 alarm rates, 5, 10, 20, and 30 alarms per hour). For each of these data sets we estimated the parameters of the queueing model from the simulation data and then computed the "theoretical" distribution of the number of busy units.

The results in each case were extraordinarily good. Plots of the theoretical frequencies and those produced by the simulation show a very close correspondence. We include here, in Figures 7 through 12, a sample of the results. We have selected for presentation data for engines and ladders in the south, north and entire Bronx at the two extreme rates of 5 and 30 alarms per hour. We should remark that although the fits are quite close "by eyeball" they fail the chi-square goodness of fit test. The reason is that the sample sizes are very large: 430 observations in the 30 alarms per hour cases and 2871 observations in the 5 alarms per hour cases. Nevertheless, we feel that these comparisons validate the model for the types of policy analyses and decision making applications for which the queueing model was created.
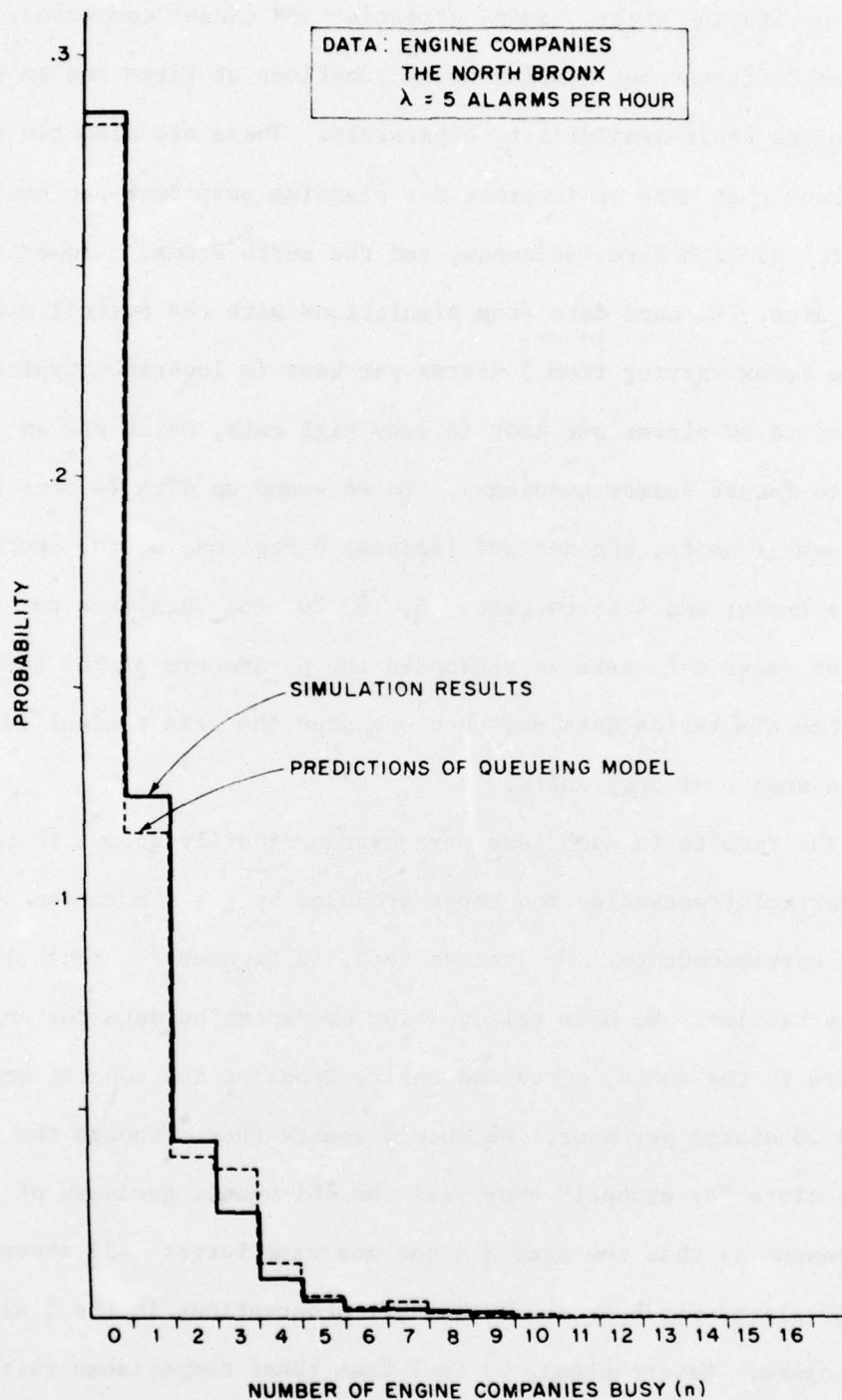
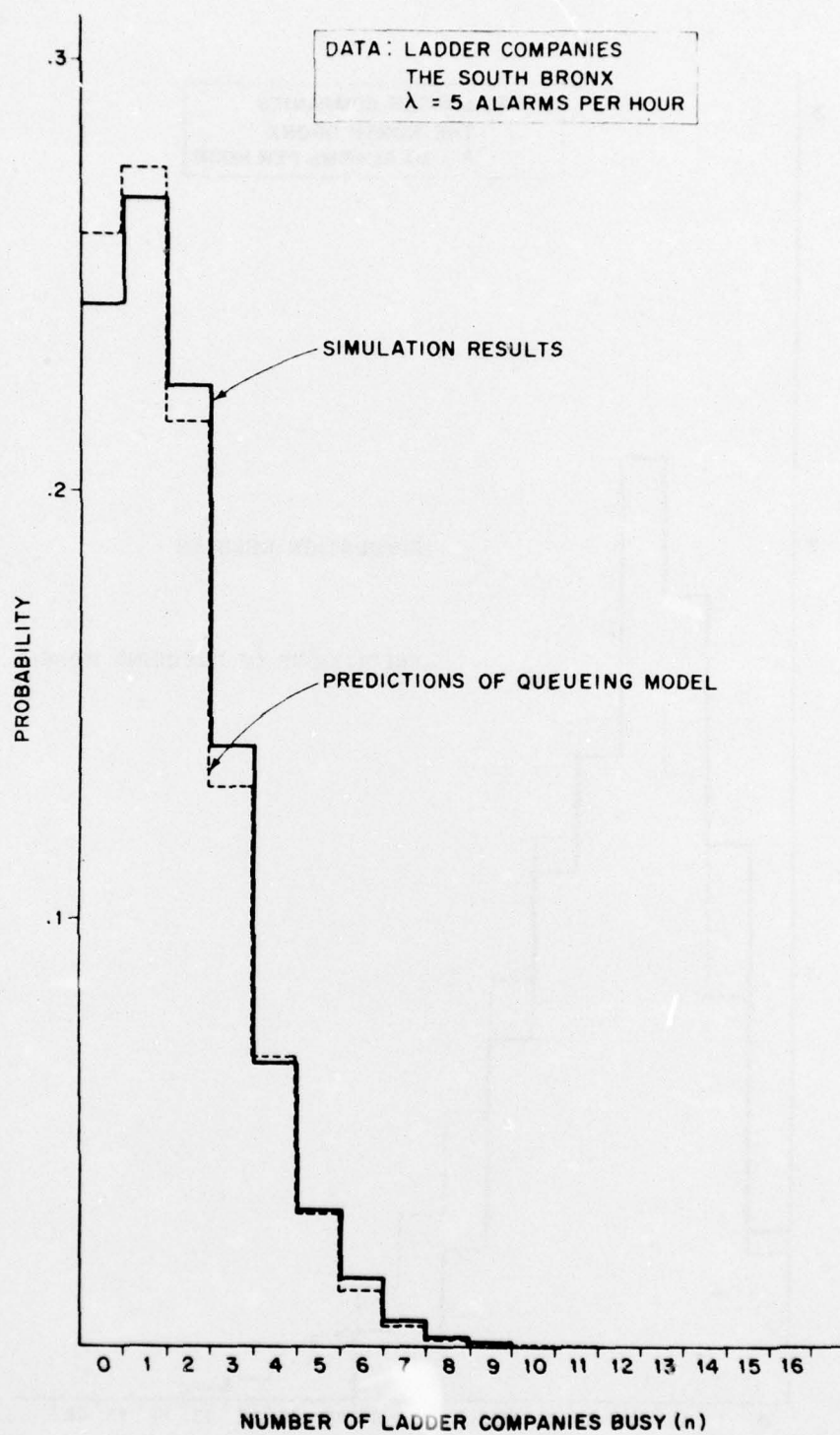Fig. 7. The probability that n engine companies are busy
(north Bronx, λ=5).

Fig. 8.   The probability that n ladder companies are busy
          (south Bronx, λ=5).
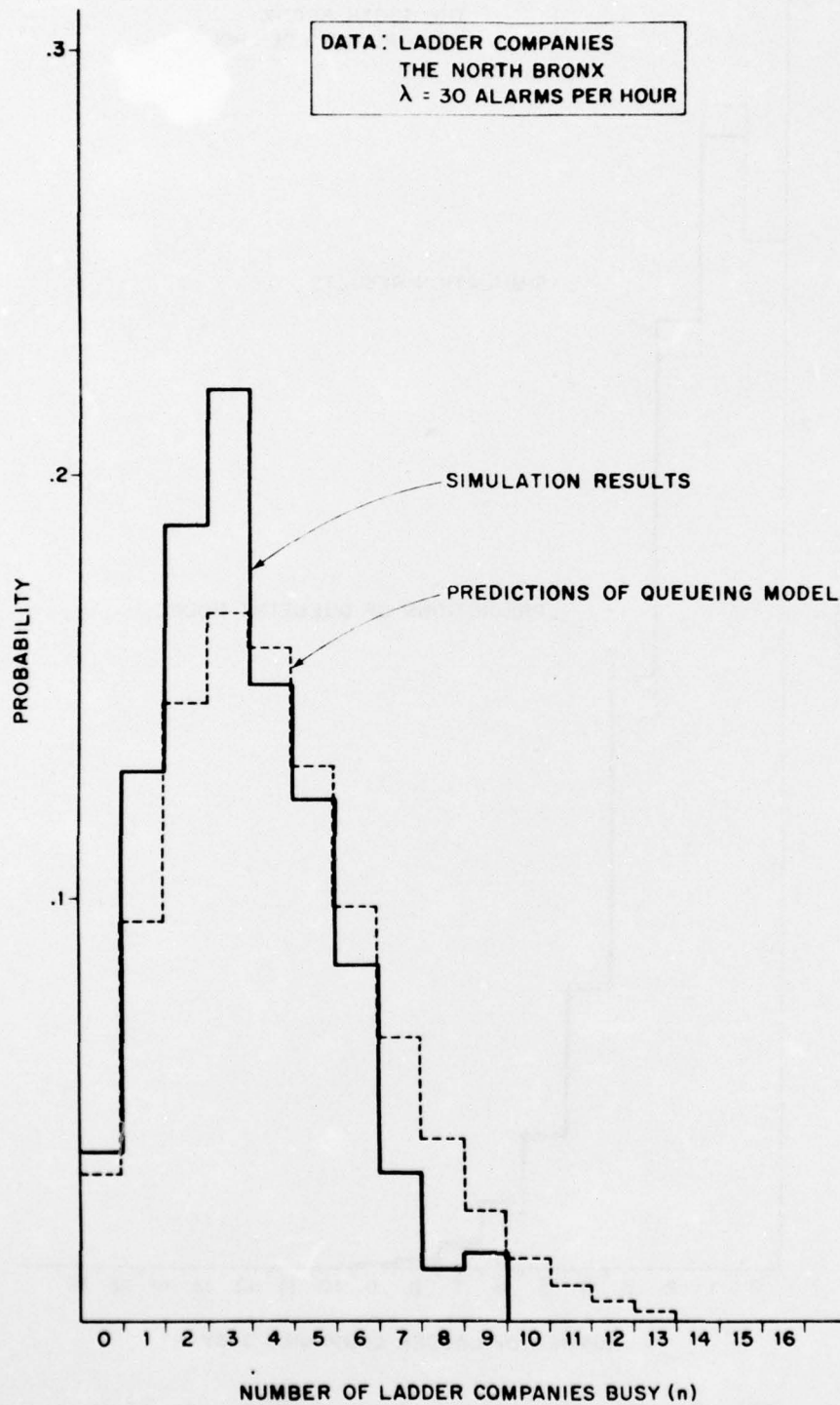
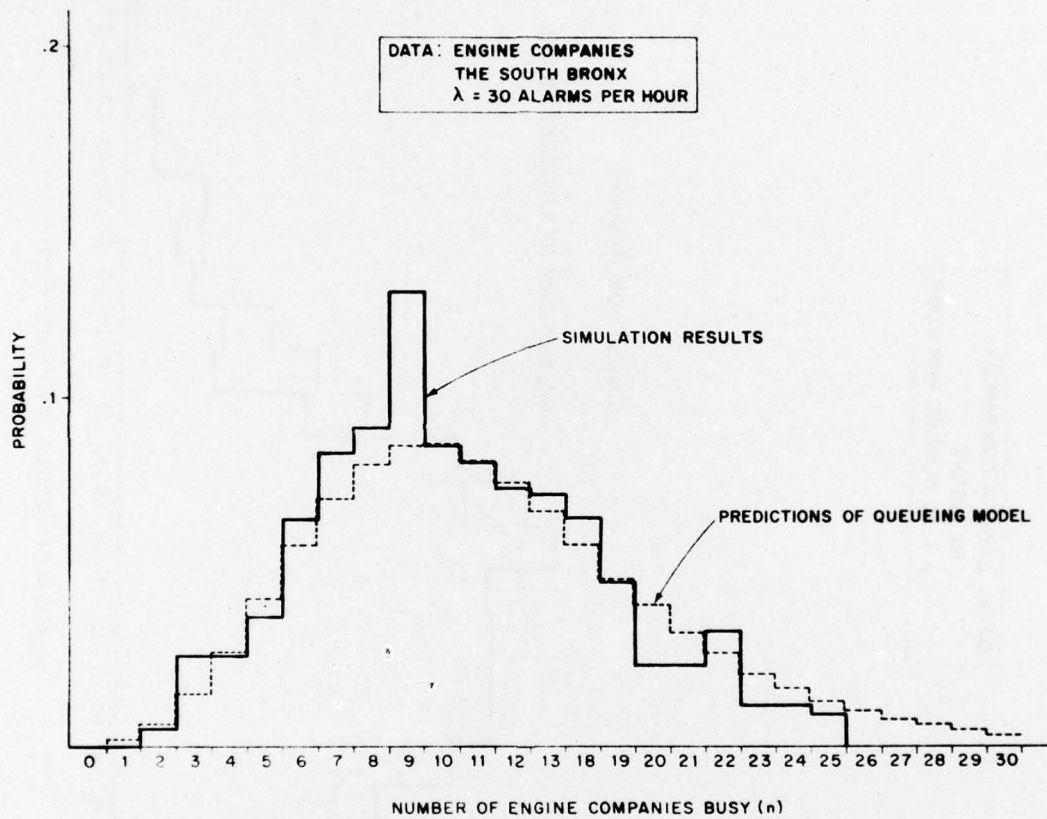Fig. 9. The probability that n ladder companies are busy (north Bronx, λ=30).

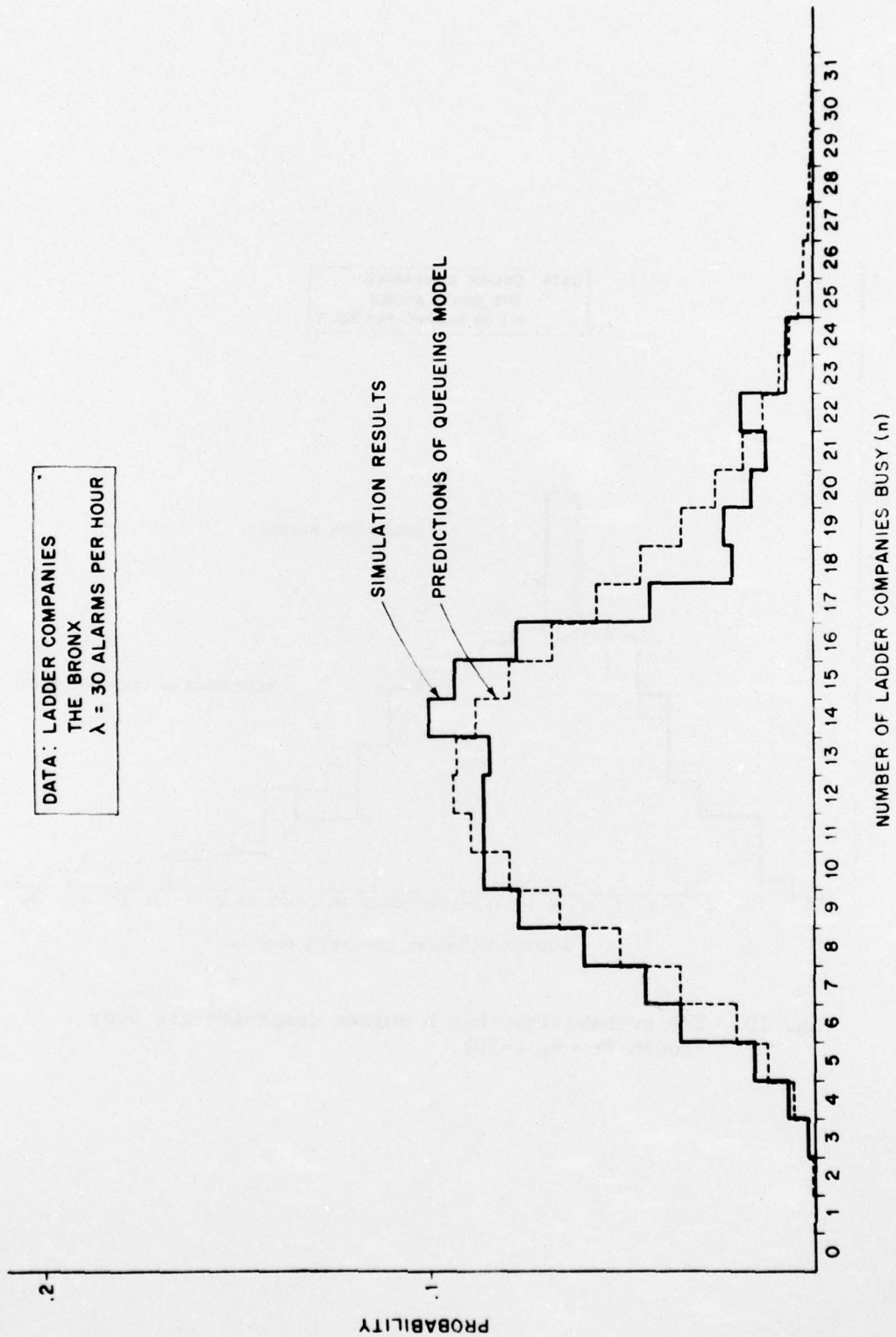Fig. 10. The probability that n engine companies are busy
(south Bronx, λ=30).

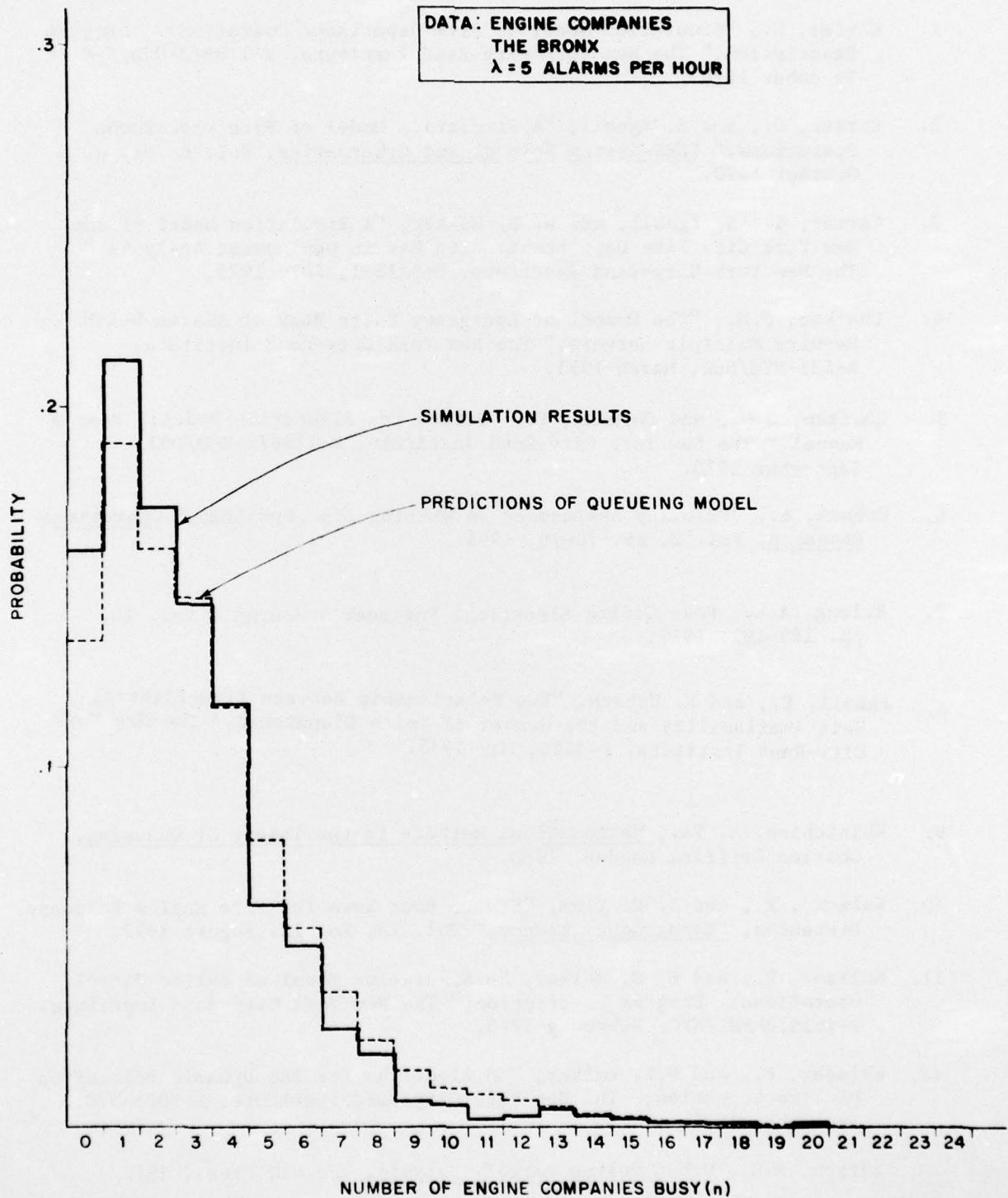Fig. 11. The probability that n ladder companies are busy (the whole Bronx, λ=30).

Fig. 12. The probability that n engine companies are busy
(the whole Bronx, λ=5).

REFERENCES

1. Carter, G., "Simulation Model of Fire Department Operations: Program Description," The New York City-Rand Institute, R-1188/2-HUD/NYC, December 1974.

2. Carter, G., and E. Ignall, "A Simulation Model of Fire Department Operations," IEEE-System Science and Cybernetics, Vol. 6, No. 4, October 1970.

3. Carter, G., E. Ignall, and W. E. Walker, "A Simulation Model of the New York City Fire Department: Its Use in Deployment Analysis," The New York City-Rand Institute, P-5110-1, July 1975.

4. Chaiken, J.M., "The Number of Emergency Units Busy At Alarms Which Require Multiple Servers," The New York City-Rand Institute, R-531-NYC/HUD, March 1971.

5. Chaiken, J.M., and Dormont, P., "Patrol Car Allocation Model: User's Manual," The New York City-Rand Institute, R-1786/2-HUD/DOJ, September 1975.

6. Cobham, A., "Priority Assignment in Waiting Line Problems," Operations Research, Vol. 2, pp. 70-76, 1954.

7. Erlang, A.K., Post Office Electrical Engineer's Journal, Vol. 10, pp. 189-197, 1917.

8. Ignall, E., and R. Urbach, "The Relationship Between Fire-Fighting Unit Availability and the Number of Units Dispatched," The New York City-Rand Institute, P-5420, May 1975.

9. Khintchine, A. Ya., Mathematical Methods in the Theory of Queueing, Charles Griffin, London, 1960.

10. Kolesar, P., and E. H. Blum, "Square Root Laws for Fire Engine Response Distances, "Management Science," Vol. 19, No. 12, August 1973.

11. Kolesar, P., and W. E. Walker, "A Simulation Model of Police Patrol Operations: Program Description," The New York City-Rand Institute, R-1625/2-HUD/NYC, February 1975.

12. Kolesar, P., and W.E. Walker, "An Algorithm For The Dynamic Relocation of Fire Companies," The New York City-Rand Institute, R-1023-NYC, September 1972.

13. Larson, R.C., Urban Police Patrol Analysis, The MIT Press, 1972.

14. Van Horn, R.L., "Validation of Simulation Results," Management Science, Vol. 17, No. 5, January 1971.